

As you may be aware, I was recently provided with data from the entire class for the course in question. This includes answers from all students on the quiz. This data was provided to me after my CAPP committee meeting. I also recently learned about statistics provided to the CAPP Committee by Dr. Thewissen. I believe this constitutes new, relevant, and substantial evidence that should be considered by the Committee.

The statistics below were provided by Dr. Thewissen to the CAPP Committee. They indicate that the "probability for all 6 questions to be identical and wrong is the product of these: 0.000036". (Note: the probability should be 0.000048. I tried to validate this with Dr. Thewissen before the Executive CAPP Committee's hearing, but Dr. Thewissen was not available.)

The 0.000048 probability was a single comparison between myself (seat 29) and the other student (seat 30). Based on the one sample, it could be easily interpreted that it is 99.999952% not possible to have 6 identical wrong unless I was cheating. The comparison could be made to a 99.999952% DNA match, thus leaving absolutely no room for any doubts. Under this type of scrutiny, it felt like everything I provided to the CAPP Committee was believed by the Committee to be excuses and dishonesty.

The new substantial evidence provided below demonstrates that, when more comparisons are done, the 0.000048 figure is in the mid-range of probabilities presented in this document.

Test had 25 questions, 6 were answered Identical and wrongly by both students									
question	A	B	C	D	E	total	yellow-2	n-2	probability of same answer
1	52	47	13	16	31	159	14	157	0.089
2	29	35	4	8	84	160	27	158	0.17
6	14	41	23	13	69	160	39	158	0.25
13	6	15	119	20	0	160	6	158	0.038
22	11	22	6	53	68	160	51	158	0.32
23	125	16	12	7	0	160	123	158	0.78

Probability for all 6 to be identical and wrong is the product of these
0.000036

underline-correct answer
yellow-marked by both students

Test had 25 multiple choice questions
15 were answered correct by the students
6 were answered wrong and identical by the two students
4 were answered wrong and different by the two students

This analysis assumes that correct answers were arrived at independently.

Dr. Thewissen Statistics

	Score	Total correct	Total wrong
Seat 29 (Endres)	72	18	7
Seat 30	64	16	9

Endres/seat 29 Probability: $0.1 \times 0.18 \times 0.26 \times 0.04 \times 0.33 \times 0.78 = 0.000048$

I took Dr. Thewissen's lead in choosing the statistical approach and broadened his original analysis. All probability cases shown in this document are based on the following conditions:

- Statistical model - the same statistical model that was used by Dr. Thewissen
- Sample pool - the entire HDS class quiz performance report I received on October 27, 2015.
- Samples include the students who have 5, 6, 7, 8, 9, 10, and 11 identical wrongs. (This does not include every possible case, only the ones I could identify by hand in the time I was given).

Substantial Evidence:

- **The 84% Similarity (table 1 – table 2):**

The 84% similarity with 6 identical wrong answers was seen as improbable, and was used as the basis to confirm that I was cheating. By extending the data set from two students to the entire HDS class we can see that it is not improbable. Table 1 shows 40+ cases that have a similarity greater than or equal to 76% with 6 or more identical wrong answers. Table 2 has close to a hundred cases with similarity greater than 80%.

- **The 6 identical wrong answers I have with seat 30 (table 3):**

When the sample pool is extended from one student to six students, it shows that for each identical wrong I have with seat 30, half of the sample pool also have the same identical wrong.

- **The 0.000048 probability (table 4 – table 11):**

I found hundreds of cases for students who have 5, 6, 7, 8, 9, 10, and 11 identical wrongs with the probabilities ranging between **0.0000001** and **0.000098**. The **0.000048** probability is neither unique nor the extreme. It falls in the mid-range of probabilities listed in this document for all 6 to be identical and wrong.

The statistics used to prove my cheating:

When the above statistics are viewed with a data set expanded from only two students to the entire HDS class, it is clear that my statistics are not abnormal, extreme, or unique in any way.

There are hundreds of examples of probabilities in-between the range of 0.0000001 and 0.000098. If my probability of 0.000048 was seen as improbable, and used as evidence to confirm I was cheating, then is it logical all other students who have a probability in-between the range of 0.0000001 and 0.000098 could be accused of cheating? Saying that there is a 99.99995% chance I was cheating also means that there is a 99.9999x% to 99.99999x% chance that all those other students were cheating as well. The only difference between myself and the other students was that my fidgeting and eye-movements caused me to look suspicious.

For multiple-choice questions, each choice has two or three conditions

No possibility of cheating:

1. Knowing
2. Guessing

Possibility of cheating:

1. Knowing
2. Guessing
3. Copy Answers

Based on the isolated statistic created from only one comparison, the 0.000048 probability was interpreted as a 99.99995% chance that I was cheating. With this kind of scrutiny:

- For questions that I got wrong, if I made any changes and my answer happened to be the same as the student in seat 30, then I must have copied the answers by briefly glancing over.
- The committee never considered the possibility that I was not cheating since there is only a 0.000048% possibility that I was not cheating. Under such light, everything I said sounded dishonest and like an excuse.
- I am also a repeat M1, and thus I would probably resort to cheating in order to pass.

The facts:

- I started strong this semester, as my HDS quiz score was in the top 15%.
- My module 1 test performance was greatly impacted due to the alleged accusation two days before my exam. Under such difficult stress, my final score was 68%, and the class average was 69%.
- I did not ask for accommodations before because there was never any need. The distractions do not bother me or impact my performance. The sole reason I will request accommodation in the future is so that my fidgeting will not be misinterpreted.

Everything I stated in my previous CAPP presentation still stands true.

I have presented solid scientific evidence that the criteria of cheating for the particular test in question should be reviewed. I thank the Executive Review Committee for the opportunity to defend my position and I sincerely hope that due consideration will be given for the evidence presented.

Similarity:**Similarity greater or equal to 76% with 6 or more identical wrong answers****Table 1 - 76% or greater similarity with 6 or more identical wrongs**

student	compare	Identical correct	Identical wrong	similarity
95	133	12	10	88%
75	145	14	8	88%
117	156	13	8	84%
137	145	15	7	88%
72	117	13	7	80%
75	137	13	7	80%
117	72	13	7	80%
137	75	13	7	80%
14	60	12	7	76%
51	137	12	7	76%
60	75	12	7	76%
75	117	12	7	76%
32	113	16	6	88%
20	117	15	6	84%
22	91	15	6	84%
32	145	15	6	84%
14	72	14	6	80%
21	140	14	6	80%
28	72	14	6	80%
32	137	14	6	80%
44	80	14	6	80%
59	89	14	6	80%
73	117	14	6	80%
89	59	14	6	80%
113	145	14	6	80%
117	73	14	6	80%
117	145	14	6	80%
129	156	14	6	80%
137	32	14	6	80%

140	21	14	6	80%
145	113	14	6	80%
145	117	14	6	80%
156	129	14	6	80%
14	112	13	6	76%
14	156	13	6	76%
16	113	13	6	76%
32	44	13	6	76%
44	32	13	6	76%
50	89	13	6	76%
51	145	13	6	76%
55	85	13	6	76%
60	145	13	6	76%
111	137	13	6	76%
112	14	13	6	76%
112	113	13	6	76%
136	139	13	6	76%
144	145	13	6	76%

Table 2 – cases with 80% or higher similarity

student	compare	similarity
29	146	92%
29	130	92%
29	32	88%
29	113	88%
32	113	88%
32	146	88%
33	69	88%
75	145	88%
95	133	88%
113	146	88%
115	146	88%
130	146	88%
137	145	88%
14	92	84%
14	113	84%

20	117	84%
22	91	84%
28	146	84%
28	115	84%
29	145	84%
29	115	84%
29	43	84%
32	145	84%
74	154	84%
117	156	84%
145	146	84%
7	86	80%
14	72	80%
14	29	80%
14	146	80%
21	140	80%
28	72	80%
28	29	80%
28	130	80%
29	74	80%
29	86	80%
31	113	80%
31	143	80%
32	137	80%
32	115	80%
32	130	80%
33	86	80%
40	74	80%
40	154	80%
43	113	80%
44	80	80%
59	89	80%
59	150	80%
72	117	80%
73	117	80%
73	113	80%
73	86	80%
74	127	80%
75	137	80%
86	33	80%
86	73	80%
86	92	80%
86	105	80%

86	122	80%
86	100	80%
89	59	80%
92	113	80%
92	86	80%
92	100	80%
99	129	80%
100	92	80%
105	86	80%
113	145	80%
113	31	80%
113	73	80%
113	92	80%
113	115	80%
113	43	80%
113	130	80%
115	32	80%
115	113	80%
115	130	80%
117	72	80%
117	73	80%
117	145	80%
122	86	80%
129	156	80%
129	145	80%
129	99	80%
129	140	80%
130	28	80%
130	32	80%
130	113	80%
137	75	80%
137	32	80%
140	21	80%
140	129	80%
143	31	80%
145	113	80%
145	117	80%
145	129	80%
156	129	80%

Statistical Analysis of identical wrongs with Student at seat 30:

These are not all the cases, but just a few listed. In the sample pool of 6 students compared with the student at seat 30.

	Q1 0.10	Q2 0.18	Q6 0.26	Q12 0.19	Q13 0.04	Q16 0.04	Q20 0.23	Q22 0.33	Q23 0.78
Seat 30	d	a	b	b	a	b	d	d	a
106		x	x		x	x			x
118			x	x		x		x	x
87	x	x		x			x	x	x
Seat 29 (Endres)	x	x	x		x			x	x
135		x	x			x			x
157				x		x		x	x

Table 3 – Identical wrong answers with student at seat 30

- For question 13 and question 16, there is 4% of the class who chose the same answer. In the small sample pool of six, 5 students have at least one identical wrong of those two questions.
- Student 106 is the only one who has the same identical wrongs for both questions with the lowest number. The probability of student 106 compare to the student (seat 30) is 0.000058.
- When the sample pool is extended to more than one, it clearly shows that for each identical wrong I have with the student at seat 30, at least half of the students from the sample have the same identical wrong.
- Is it rational to use the statistics of student 106 to say that he/she might be cheating by somehow copying the answers from student at seat 30? If not, then how can my probability with seat 30 be used to prove I was cheating? The only additional factor is my fidgeting.

HDS Quiz Probability samples for all 5 to be identical and wrong:**Probability range for all 5 identical and wrong: 0.000003 - 0.000068**

Student	Compare to	Probability	Same wrong Answers
18	49	0.000077	5
18	108	0.000016	5
18	141	0.000068	5
(seat 30)	340*	0.000058	5
24	65	0.000009	5
24	67	0.000003	5
108	82	0.000038	5
106	22	0.000058	5
141	18	0.000068	5

Table 4 - Probability with 5 identical wrong answers**HDS Quiz Probability samples for all 6 to be identical and wrong:****Probability range: 0.000001 - 0.000098**

Student	Compare to	Probability	Same Wrong Answers
11	83	0.000033	6
11	108	0.000016	6
11	152	0.000048	6
12	61	0.000088	6
13	24	0.000011	6
13	27	0.000007	6
13	47	0.000001	6
23	83	0.000031	6
34	27	0.000035	6
34	61	0.000092	6
42	67	0.000037	6
42	90	0.000069	6
46	153	0.000060	6
60	102	0.000070	6
61	12	0.000088	6
61	102	0.000016	6
61	132	0.000056	6
67	109	0.000079	6
67	153	0.000065	6

68	13	0.000050	6
68	132	0.000056	6
91 (Endres)	22 (seat 30)	0.000048	6
96	161	0.000068	6
103	16	0.000071	6
108	18	0.000016	5
108	141	0.000026	6
132	110	0.000031	6
135	42	0.000067	6
149	141	0.000067	6
153	41	0.000055	6
153	114	0.000098	6
155	161	0.000077	6
161	153	0.000030	6
69	118	0.000089	6
31	61	0.000087	6
104	46	0.000090	6
70	42	0.000058	6
104	46	0.000090	6
111	23	0.000089	6
90	42	0.000069	6
94	61	0.000055	6
139	136	0.000051	6

Table 5 – Probability with 6 identical wrongs

HDS Quiz Probability samples for all 7 to be identical and wrong:**Probability range: 0.000004 – 0.000098**

Student	Compare to	Probability	Same Wrong Answers
5	19	0.000073	7
5	46	0.000067	7
5	83	0.000004	7
5	120	0.000088	7
5	126	0.000078	7
8	37	0.000025	7
11	138	0.000052	7
13	52	0.000082	7
16	121	0.000047	7
24	16	0.000009	7
27	68	0.000004	7
27	132	0.000010	7
34	83	0.000003	7
37	157	0.000069	7
41	42	0.000071	7
41	67	0.000033	7
41	83	0.000024	7
41	136	0.000044	7
42	49	0.000005	7
42	83	0.000019	7
47	34	0.000014	7
47	67	0.000005	7
48	161	0.000005	7
49	30	0.000040	7
49	61	0.000011	7
49	62	0.000020	7
49	110	0.000026	7
49	138	0.000076	7
50	87	0.000077	7
60	121	0.000049	7
79	147	0.000098	7
82	110	0.000010	7
88	110	0.000004	7
106	118	0.000022	7
109	119	0.000059	7
110	34	0.000012	7
110	138	0.000097	7
114	123	0.000031	7
121	103	0.000080	7
157	119	0.000010	7
158	147	0.000083	7
85	42	0.000017	7
101	46	0.000071	7
116	34	0.000043	7

Table 6 – Probability with 7 identical wrong answers

HDS Quiz Probability samples for all 8 to be identical and wrong:**Probability range: 0.0000002 - 0.000096**

Student	Compare to	Probability	Wrong Answers
9	37	0.000071	8
12	120	0.000068	8
19	46	0.000074	8
25	49	0.000015	8
25	78	0.000023	8
25	82	0.000096	8
25	158	0.000061	8
27	47	0.0000002	8
34	126	0.000026	8
34	142	0.000015	8
36	119	0.000023	8
37	121	0.000039	8
41	64	0.000057	8
41	90	0.000023	8
42	25	0.000030	8
46	48	0.000022	8
46	126	0.000063	8
48	87	0.000048	8
48	110	0.000042	8
48	155	0.000016	8
50	119	0.000029	8
50	120	0.000061	8
57	118	0.000059	8
57	157	0.000059	8
61	83	0.000001	8
78	42	0.000006	8
87	138	0.000009	8
88	155	0.000040	8
110	114	0.000004	8
110	125	0.000011	8
110	126	0.000013	8
118	34	0.000030	8
118	93	0.000022	8
118	142	0.000084	8
118	144	0.000035	8
120	125	0.000069	8

Table 7 – Probability with 8 identical wrong answers

HDS Quiz Probability samples for all 9 to be identical and wrong:**Probability Range: 0.000001 - 0.000092**

Student	Compare to	Probability	Wrong Answers
3	87	0.000005	9
12	25	0.000034	9
23	11	0.000001	9
25	37	0.000017	9
48	97	0.000092	9
49	82	0.000001	9
110	46	0.000002	9
120	25	0.000030	9

Table 8 – Probability with 9 identical wrong answers**HDS Quiz Probability samples for all 10 to be identical and wrong:****Probability Range: 0.000001 - 0.000019**

Student	Compare to	Probability	Wrong Answers
9	25	0.000004	10
25	118	0.000019	10
133	95	0.000001	10

Table 9 – Probability with 10 identical wrong answers**HDS Quiz Probability samples for all 11 to be identical and wrong:****Probability Range: 0.0000004**

Student	Compare to	Probability	Wrong Answers
157	118	0.0000004	11

Table 10 – Probability with 11 identical wrong answers

Timestamps

The timestamps from the laptops and the video cannot be synced up. The video showed the student in seat 30 pass behind me at 29:15, but his timestamps show that he quit the exam at 34:42. There were less than 4% of the timestamps that show we were overlapping. However, the following 20-second timestamp sequence is exactly identical. It does not seem possible because it would require me to predict when he switches questions:

Student in seat 30

6	4 Choice	8:30:33 AM	Navigating	Choice(s): B
7	3 Choice	8:30:38 AM	Navigating	Choice(s): D
8	3 Choice	8:30:42 AM	Navigating	Choice(s): B
9	3 Choice	8:30:46 AM	Navigating	Choice(s): A
10	3 Choice	8:30:49 AM	Navigating	Choice(s): B
11	3 Choice	8:30:53 AM	Navigating	Choice(s): B

Student in seat 29 (Julian)

6	5 Choice	8:30:33 AM	Navigating	Choice(s): B
7	6 Choice	8:30:38 AM	Navigating	Choice(s): D
8	6 Choice	8:30:42 AM	Navigating	Choice(s): B
9	6 Choice	8:30:46 AM	Navigating	Choice(s): A
10	9 Choice	8:30:49 AM	Navigating	Choice(s): B
11	9 Choice	8:30:53 AM	Navigating	Choice(s): B

	Student in seat 29 (Endres)	Student in seat 30
Average revisits per question	10	4.5